

Minireview

Coordinated gene expression by post-transcriptional regulons in African trypanosomes

Marc Ouellette and Barbara Papadopoulou

Addresses: Centre de Recherche en Infectiologie and Département de Microbiologie-Infectiologie et Immunologie, Université Laval, Québec, G1V 4G2, Canada.

Correspondence: Marc Ouellette. Email: Marc.Ouellette@crchul.ulaval.ca. Barbara Papadopoulou. Email: barbara.papadopoulou@crchul.ulaval.ca.**Abstract**

The regulation of gene expression in trypanosomes is unique. In the absence of transcriptional control at the level of initiation, a subset of *Trypanosoma brucei* genes form post-transcriptional regulons in which mRNAs are co-regulated in response to differentiation signals.

See research articles <http://www.biomedcentral.com/1471-2164/10/427>, <http://www.biomedcentral.com/1471-2164/10/482> and <http://www.biomedcentral.com/1471-2164/10/495>.

The kinetoplastid parasites diverged early in the eukaryotic branch of life and several of their members are responsible for some of the great scourges of humanity, including sleeping sickness (caused by *Trypanosoma brucei*), Chagas disease (caused by *Trypanosoma cruzi*) and leishmaniasis (caused by *Leishmania* species). These parasites are distinguished by the kinetoplast, the dense DNA-containing region inside the single large mitochondrion. Because of their medical and veterinary importance, these parasites have been intensively investigated and their study has led to the discovery of a number of novel basic mechanisms, including *trans*-splicing, RNA editing, glycosylphosphatidylinositol-anchoring of membrane proteins, and the polarization of T-cell subsets in immunology. The regulation of gene expression in these early-diverging eukaryotes displays some unique features. The findings of three papers published recently in *BMC Genomics* [1-3] show that despite a lack of transcriptional control at the level of initiation, the expression of subsets of genes in *T. brucei* is regulated during differentiation in a coordinated fashion at the post-transcriptional level. This leads to 'post-transcriptional regulons', a phenomenon recently recognized in many organisms (reviewed in [4]) and proposed to exist in *T. brucei* [5,6].

Constitutive RNA polymerase-II-mediated transcription in kinetoplastids

The 'TriTryp' (*Leishmania* species, *T. brucei* and *T. cruzi*) genomes are organized into large gene clusters that are constitutively co-transcribed by RNA polymerase II (Pol II)

to yield polycistronic pre-mRNAs - that is, RNA containing multiple protein-coding sequences [7]. In contrast to the DNA operons of prokaryotes, however, there is no evidence of functional clustering within these polycistronic transcription units.

These polycistronic pre-mRNAs are processed by two coupled cleavage reactions - a *trans*-splicing reaction that adds a capped spliced leader RNA of 39 nucleotides to the 5' terminus of all the known protein-coding RNAs, and 3'-polyadenylation (Figure 1). This unusual mechanism of generating mature mRNAs precludes individual regulation of gene expression at the level of initiation of transcription. Pol II promoters are indeed elusive in these parasites and sequence analysis has revealed a paucity of the basal Pol II transcription factors in their genomes [7].

The regions between polycistronic units are known as strand-switch regions (SSRs). Depending on the transcriptional orientation, the units can be convergent (transcriptional operons on opposite strands are converging towards the SSRs) or divergent (transcriptional operons start on opposite strands of the SSRs and diverge from one another) (Figure 1). SSRs associated with divergent units in *Leishmania* have been shown to be preferential sites of transcription initiation, whereas convergent SSRs were enriched for transcription termination sites [8]. Recent chromatin immunoprecipitation and sequencing (ChIP-seq) experiments examining the genome-wide distribution of chromatin components in *T. brucei* showed that the seemingly unregulated transcription of trypanosomes is directed by histone post-translational modifications, thus indicating the important role that chromatin modifications play in polycistronic transcription initiation and termination [9]. While divergent SSRs were indeed found to be potential transcription start sites, many other start sites were also pinpointed, often downstream of tRNA genes [9] (Figure 1). While we refrain from putting *T. brucei* and *Leishmania* under the same regulatory umbrella, it is intriguing to note that histone modifications were also found in divergent SSRs in *Leishmania* [10],

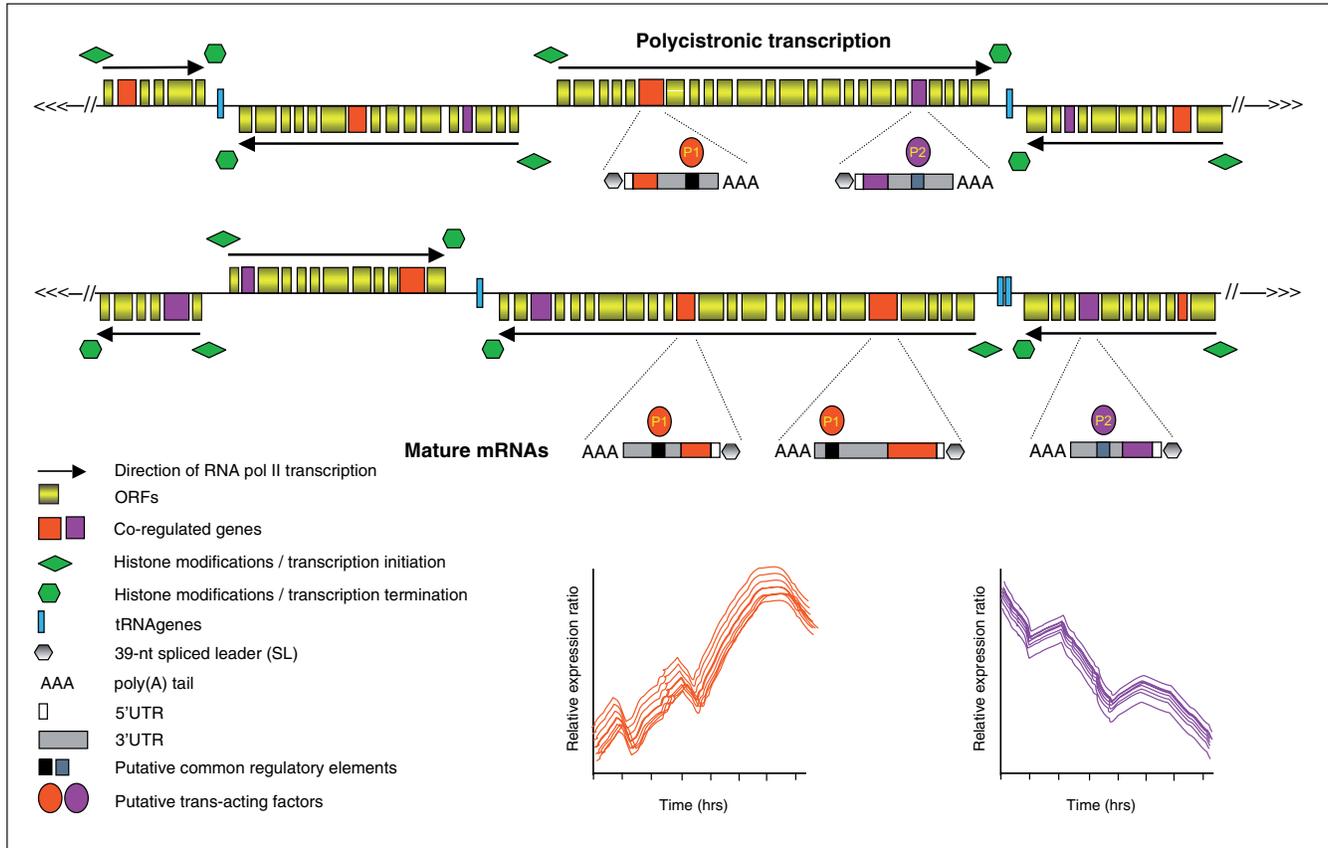


Figure 1

Coordinated post-transcriptional regulation of *T. brucei* mRNAs during differentiation. Schematic diagram of putative regions of two *T. brucei* chromosomes. Genes in *T. brucei* are organized into long polycistronic clusters that are co-transcribed by RNA polymerase II (Pol II) to yield polycistronic pre-mRNAs, which are processed by *trans*-splicing (addition of a capped spliced leader RNA of 39 nucleotides to the 5' terminus of transcripts) and 3'-polyadenylation to generate mature mRNAs. Transcription initiates from divergent strand-switch regions (SSRs) and terminates at convergent SSRs, where tRNA genes are often located (although they can be present at non-SSRs). Initiation and termination of transcription in *T. brucei* are characterized by distinct chromatin variants and modifications [9]. Three recent reports [1-3] indicate that subsets of trypanosome genes form post-transcriptional regulons during *T. brucei* life-cycle transitions. Two hypothetical post-transcriptional regulons formed during differentiation are shown. Subsets of genes (here shown in either orange or violet) have common regulatory elements or conserved secondary structures within the 3'UTRs. These are recognized by *trans*-acting factors (specific for either the set of genes in orange or in violet, and either stabilizing or destabilizing mRNAs), which allow a coordinated regulation of sets of mRNAs. This is illustrated in the two lower graphs, where mRNA levels are plotted against the differentiation process with time. The mRNA levels of the cluster of genes in orange are increasing coordinately upon differentiation, whereas the cluster of genes in violet are decreasing upon differentiation in a coordinated fashion.

although additional sites outside SSRs were also identified. Altogether, these findings support the view that transcription in kinetoplastid parasites is constitutive and that chromatin structure, in part mediated through histone modifications, will determine transcription start and termination sites. These do not seem to be sequence-specific and several of these sites (but clearly not all) are within SSRs.

Post-transcriptional control of gene expression in kinetoplastids

Kinetoplastid parasites have relatively complex life cycles during which they undergo extensive developmental changes.

T. brucei cycles between the bloodstream of mammalian hosts and the tsetse fly vector. This cycling is accompanied by changes in morphology, in metabolism, and in RNA and protein expression. Because the genome of *T. brucei* is transcribed mostly constitutively, as previously described, regulation of gene expression occurs almost exclusively by post-transcriptional mechanisms. These include mRNA processing, mRNA degradation and translational efficiency, and protein processing, modification and stability [11]. Several studies have reported that sequences within 3'-untranslated regions (3'UTRs) play a key role in controlling either the stability of kinetoplastid mRNAs or the efficiency of their translation [11].

Within the mammalian bloodstream, trypanosomes grow as long slender forms. When parasitemia reaches a threshold, trypanosomes transform into a quiescent short stumpy form. Within the tsetse fly vector, this quiescent form rapidly transforms into procyclic parasites in the insect midgut. These transform further into epimastigote and metacyclic forms within the insect. The three recent *BMC Genomics* papers by Kabani and colleagues [1], Jensen and colleagues [2], and Queiroz and colleagues [3] have taken advantage of whole-genome oligonucleotide-based DNA microarrays to study the changes in mRNA levels during the important *T. brucei* life-cycle transitions from long and slender to short and stumpy, and thereafter from stumpy to tsetse-midgut procyclics [1-3].

Previous transcriptomic analyses revealed that only a small proportion (2 to 5%) of mRNAs were modulated throughout the life cycle of *T. brucei*, and that this paralleled observations in the related *Leishmania* (reviewed in [11]). However, the data reported by Jensen and colleagues [2] now suggest that expression of up to 25% of the coding RNAs varies in at least one part of the parasite's life cycle. These numbers are clearly higher than earlier reports, although significant variation was observed among the three new studies [1-3]. These variations could partly be accounted by the fold-threshold changes used as a criterion for change, as the studies by Jensen *et al.* [2] and Queiroz *et al.* [3], which retained smaller-fold change criteria, found greater numbers of differentially expressed genes. It remains to be determined whether small changes in mRNA levels will impact on protein production and activity, but this new work [1-3] gives eloquent examples of changes in mRNA levels correlated with changes in protein or metabolite levels. Even more remarkable is the observation that the expression of several of the differentially expressed genes was modulated post-transcriptionally in a coordinated fashion.

Post-transcriptional regulons

Post-transcriptional mechanisms of regulation can influence splicing, transport, stability, localization, and translation of messenger RNAs. This post-transcriptional regulation is mediated by *trans*-acting factors (proteins, RNAs and metabolites) that recognize *cis*-acting sequences or structures, usually within the 3'UTRs of mRNAs. If a protein were to recognize a group of mRNAs containing the same sequences in their 3'UTRs, hence modulating the stability of this group of mRNAs in a coordinated fashion, it would lead to a post-transcriptional regulon (reviewed in [4]). Post-transcriptional regulons have been described in budding yeast, fruit fly and mammalian cells [4], and possibly the best-studied examples are the Pumilio RNA-binding protein family members (PUFs) in yeast. Indeed, each yeast PUF was found to bind and destabilize a distinct subset of mRNAs coding for proteins with related functions [12]. As kinetoplasts rely exclusively on post-transcriptional

mechanisms, post-transcriptional regulons are likely candidates for gene regulation in these parasites. Recent studies have indeed provided evidence for this concept [5,6] and the three *BMC Genomics* papers [1-3] show the potential for many additional putative post-transcriptional regulons in *T. brucei*.

These new discoveries were rendered possible by a number of technological improvements (DNA microarrays and stringent statistical analyses) and more sophisticated experimental design (involving defined parasite genetic lines, larger numbers of biological replicates, and careful monitoring of the time course of parasite differentiation). The level of co-regulation of some *T. brucei* genes was striking and several clusters of coordinated gene expression were highlighted. Most clusters contained genes with a variety of functions, although some co-regulated genes were functionally related. These observations further supported the notion that despite an absence of control of transcriptional initiation, gene expression can be finely tuned through post-transcriptional mechanisms during the *T. brucei* life cycle. Several of the co-regulated clusters were logical and consistent with the biology of the parasite. Indeed, despite non-identical experimental set-ups between the three studies [1-3], a number of common observations were made (although admittedly, many differences were also apparent). For example, the RNAs of genes coding for proteins involved in the translational machinery were coordinately downregulated during the transition from long slender to short stumpy bloodstream forms, but their expression increased *en bloc* on transformation from short stumpy to procyclics [1-3]. Within some of the clusters, there were many genes of unknown function co-regulated with genes of known function. This clustering can lead to testable hypotheses for examining the role of hypothetical genes.

Regulatory factors of post-transcriptional regulons

Post-transcriptional regulation of gene expression networks is a ribonucleoprotein-driven process, in which the levels of subsets of mRNAs are coordinately regulated, primarily by *trans*-acting factors. These factors interact with regulatory elements that are shared between the co-regulated mRNAs (Figure 1). Searches for shared motifs in clusters of co-regulated genes in *T. brucei* met with limited success [2,3], with the exception of the transcripts upregulated in stumpy forms, which were greatly enriched for a hexamer sequence 150 nucleotides downstream of the translation stop codon [1]. The role of this sequence awaits further experimental testing, but if it is involved in coordinated gene expression, it could be used to isolate the putative *trans*-acting factors. One such factor, PUF9, was recently isolated along with its putative *cis*-acting sequence, a heptamer contained in the 3'UTRs of several *T. brucei* mRNAs [6]. PUF9 was shown to stabilize targeted mRNAs in the S-phase of the cell cycle, and these mRNAs would

constitute a post-transcriptional regulon involved in the replicative process.

Interestingly, genes encoding RNA-binding proteins were often found in the clusters of co-regulated genes and, as suggested in [3], some of these proteins might regulate the expression of genes that are part of the regulon. In contrast to *Leishmania* species and *T. cruzi*, RNA interference (RNAi) functions well in *T. brucei*, and this technique can be used at a genome-wide scale. By silencing genes coding for putative RNA-binding proteins and using microarrays to look for post-transcriptional regulons during differentiation (or other biological processes), it should be possible to isolate *trans*-acting factors involved in post-transcriptional control of gene expression. Genome analysis has revealed that kinetoplastid parasites have an unusually large repertoire of genes coding for RNA-binding proteins [7], which is consistent with organisms relying on post-transcriptional mechanisms for gene regulation.

Control of gene expression in kinetoplastid parasites is unique, and relies exclusively on post-transcriptional mechanisms. Recent papers have now indicated that in *T. brucei* differentiation, some of the regulation is highly coordinated. Genes involved in processes other than differentiation might possibly also be regulated by coordinated RNA stability, as shown for the *T. brucei* replication process [6]. It is also likely that the regulation of many other genes will be at the translational or post-translational level. Trypanosomes and the related *Leishmania* species depend on the dynamics of gene expression to regulate differentiation, adaptation to stress, and proliferation in response to diverse environmental signals within different hosts.

It remains to be seen whether *T. cruzi* and *Leishmania* species, whose genomes are highly syntenic (that is, similar in the order of the genes) with *T. brucei* [7], will use similar strategies for regulating mRNA levels. A recent transcriptomic analysis has shown that about 50% of *T. cruzi* genes are differentially expressed during development [13], but several reports from *Leishmania* did not suggest such extensive changes (reviewed in [11]). Recent evidence, however, would suggest that many *Leishmania* genes are regulated post-transcriptionally by small degenerate inactive retroposons (SIDER1 and SIDER2) in their 3'UTRs (reviewed in [11]).

Kinetoplastid parasites have a proven record in generating novel concepts involved in the regulation of gene expression. The quasi-exclusive dependence on post-transcriptional mechanisms for coordinated gene expression makes *T. brucei* an interesting model system for deciphering mechanisms governing the generation of post-transcriptional regulons. In the mid-term, this work may also

lead to novel urgently required therapeutic targets and strategies for controlling important human diseases caused by these deadly parasites.

Acknowledgements

Work in the labs of MO and BP is funded by the Canadian Institutes of Health Research (CIHR). MO and BO are Burroughs Wellcome Fund investigators and MO is the holder of a Canada Research Chair.

References

1. Kabani S, Fenn K, Ross A, Ivens A, Smith TK, Ghazal P, Matthews K: **Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in *Trypanosoma brucei*.** *BMC Genomics* 2009, **10**:427.
2. Jensen BC, Sivam D, Kifer CT, Myler PJ, Parsons M: **Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*.** *BMC Genomics* 2009, **10**:482.
3. Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C: **Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons.** *BMC Genomics* 2009, **10**:495.
4. Keene JD: **RNA regulons: coordination of post-transcriptional events.** *Nat Rev Genet* 2007, **8**:533-543.
5. Estevez AM: **The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes.** *Nucleic Acids Res* 2008, **36**:4573-4586.
6. Archer SK, Luu VD, de Queiroz RA, Brems S, Clayton C: ***Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle.** *PLoS Pathog* 2009, **5**:e1000565.
7. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchetti G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, et al.: **The genome of the kinetoplastid parasite, *Leishmania major*.** *Science* 2005, **309**:436-442.
8. Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ: **Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region.** *Mol Cell* 2003, **11**:1291-1299.
9. Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyó D, Wang X, Dewell S, Cross GA: **Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*.** *Genes Dev* 2009, **23**:1063-1076.
10. Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ: **Histone acetylations mark origins of polycistronic transcription in *Leishmania major*.** *BMC Genomics* 2009, **10**:152.
11. Haile S, Papadopoulou B: **Developmental regulation of gene expression in trypanosomatid parasitic protozoa.** *Curr Opin Microbiol* 2007, **10**:569-577.
12. Gerber AP, Herschlag D, Brown PO: **Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast.** *PLoS Biol* 2004, **2**:E79.
13. Minning TA, Weatherly DB, Atwood J 3rd, Orlando R, Tarleton RL: **The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*.** *BMC Genomics* 2009, **10**:370.

Published: 14 December 2009

doi:10.1186/jbiol203

© 2009 BioMed Central Ltd